International Journal of Computer Science Engineering
and Information Technology Research (IJCSEITR)
ISSN (P): 2249–6831; ISSN (E): 2249–7943
Vol. 12, Issue 1, Jun 2022, 71–82

COMPUTATION OF AMAZON STOCK PRICE FOR THE NEXT DECADE USING ARIMA AND SARIMAX MODEL

HET TRIVEDI, MANSI GOPANI, HARSH LOTIA, DRUMIL JOSHI & RANJUSHREE PAL

Electronics and Telecommunication Engineering, Dwarkadas J. Sanghvi College of Engineering, Maharashtra, India

ABSTRACT

Company shares are probably the most popular financial tool built to build wealth and are the basis of any investment portfolio. Advances in trading technology have opened markets so that today almost anyone can own shares. In the last few decades, there has been a marked increase in the general interest in the stock market. In a volatile financial market, such as the stock market, it is important to have a more accurate forecasting of future trends. Due to financial constraints and the benefit of recording, it is mandatory to have a secure stock price forecast. In this research paper, we compare two models to predict Amazon company's stock market trends based on technical analysis using stock market history data. Two statistical models are ARIMA and SARIMAX. This will change the process of stock price indexing in the future and aid financial professionals to select the best time to buy and / or sell stocks. Results are displayed visually using Python editing language and Microsoft Excel. The results obtained reveal that the SARIMAX model has the potential to predict short-term stock market trends.

KEYWORDS: Amazon Stock Price, Arima & Sarimax Model

Received: Jan 01, 2022; **Accepted**: Jan 31, 2022; **Published**: Feb 08, 2022; **Paper Id**: IJCSEITRJUN20228

Original Article

1. INTRODUCTION

Everyone wants to become rich in his/her life with less effort and maximum advantages. The question arises how it can be done. In today's world there has been an increase in the number of stock market traders, one of its reasons being that people consider it as an instrument for earning quick wealth. The Stock Market has become a very popular tool for investment, where people buy and sell stocks to earn huge profits. But with becoming wealthier it also brings together the risks involved with it. So, people always try to find different ways to reduce the risk factor and increase their profits. In a financially explosive and volatile market it would be great if the movement of the future stock indices could be predicted beforehand which would help the investors invest their money in correct directions and help them avoid wastage of money. Forecasting the stock data based on the past data has become very important to understand the market trends. There have been multiple attempts to predict future market trends with the help of different Machine Learning models and Artificial Neural Network (ANNs) models. Closed time forecasting plays a main role in finance and economics. Financial time series data, especially stock market data is very hard in decomposition and forecasting because the data are non-linear and non-stationary with high heteroscedasticity. In this research ARIMA - Auto Regressive Integrated Moving Average which is an old model but still has its many applications in the field of financial time series is used. The ARIMA model is a statistical tool for analysing and forecasting time series data by modelling the data's associations. This model excels in making short-term forecasts. Moreover, it uses only the past stock market data for generalising and predicting the future trends, so the number of parameters required is very less which helps in improving its accuracy. The theoretical

ARIMA model and structural linkages, on the other hand, are not distinguishable from several other basic forecasting models. In this article the researchers are applying the ARIMA model for short term market data on Amazon stocks based on its past data. The results obtained from this can be applied for short term forecasting and assist in investment decision making.

2. LITERATURE SURVEY

In [1] they have explained about the future predictions about different stocks with the help of different machine learning algorithms. In today's world there has been an increase in the number of stock market traders, one of its reasons being that people consider it as an instrument for earning quick wealth. It has become a very popular tool for investment, where people buy and sell stocks to earn huge profits. Everyone wants to be rich without many efforts and easy returns, so trading in the stock market can earn one huge profit but it is very risky. If the stock prices can be predicted beforehand then it becomes easier where to invest and helps in avoiding wastage of money. For stock market prediction in [1] they have implemented different algorithms such as Linear Regression (LR), Three Months Moving Average (3MMA) and Exponential Smoothing on different stocks past data and trends. After applying these machine learning algorithms hypothesis of these was obtained that exponential smoothing prediction resulted in less error and greater accuracy. So, it is considered as the best stock market predictor with general trend analysis among these three algorithms. Time Series Forecasting method was also used to predict stock prices for next month. Applying these methodologies future stock market trends prediction became easier.

Time series data forecasting, especially in the field of finance and economics plays a main role. Financial data such as stock market data which is highly linear and highly stationary is very hard to decompose and predict. So, this has led to many new models being discovered but ARIMA which is an old model is still widely used for such applications. ARIMA model is a statistical method which is very useful for short term prediction as the number of parameters required is very less. This model is a combination of three other models which are an autoregressive (AR) model and a moving average, (MA) model, {et}: white noise (WN) process. The MINTAB software is used to get the results wherein the dataset is from the Amman Stock Exchange (ASE). In [2] RMSE is selected as the criteria as the fitted ARIMA model has less RMSE (around 4.00). The results obtained with the best ARIMA model (2, 1, 1) while other models had higher values. So, the output of this research was that the ARIMA model works reasonably well with short term forecasting.

Banking time series data is generally very hard to predict and decompose. This has led to much research on this topic and resulted in many fit models being introduced in forecasting accuracy. Banking data from Amman Stock Exchange (ASE) in Jordan was selected as a tool in [3]. The ARIMA model is used for such banking stock market data because the data is non-stationary and non-linear with high heteroscedasticity. The auto-regressive moving average (ARMA) model contains three combination models which are: autoregressive (AR) model and a moving average, (MA) model, and {et}: white noise (WN) process. Three types of accuracy criteria had been adopted to compare the performance of the model which are Mean square error (MSE), Root mean squared error (RMSE) and mean absolute error (MAE). Eventually the best fitted model with less RMSE criteria was selected. The results on real banking stock market data showed that the ARIMA model is good for short term prediction but the forecasting accuracy of the ARIMA model diminishes gradually from period to period.

The art of trading in the stock market to earn profits is not that easy. To attain this objective much research has been made to predict and forecast the future trends of a stock. To know how the stock would perform in future its present

scenarios are looked upon and analysed. In this research [4] data mining techniques are used to develop the prediction model and R programming language is used to visualise the results. The ARIMA model is used for predicting stock patterns based on its previous trends on pre-processed data. Data visualisation is done using R for short term investment assistance. Correlations can be obtained on visualisation of the results to get the predictions. A prediction model forecasting stock market trends using the time series data gave the results in favour of the ARIMA model that was useful in predicting stock indices on a short-term basis which could guide the investors in the stock market to make profitable investment decisions.

Predicting stock prices has always attracted interest of many investors because of its financial and economic benefits it provides. In this survey [5] the researchers have worked on improving the accuracy of the ARIMA model, wherein a study had been conducted on National Stock Exchange (NSE) based fifty-six Indian stocks from seven different sectors. Akaike Information Criterion has been used for the comparison and parameterization of the model. Further to check whether the model is appropriate ACF, and Partial Autocorrelation Function (PACF) is used by identifying the lag in the model. After prediction the Mean Absolute Error (MAE) method was used to measure the accuracy. After applying the model for all sectors, accuracy in predicting the stock prices was found to be above 85%, which indicated that the ARIMA model gave good accuracy.

3. METHODOLOGY

3.1 Decomposition of Series

Time series data can exhibit a wide range of patterns, so Decomposition is a quantitative approach that categorises historical data into multiple components each signifying a fundamental pattern category and it further uses them to provide a more accurate forecast by identifying seasonality and trend from a series data. Seasonality is a time series component in which the data undergoes regular and predictable changes that occur at specified cycles less than a year, such as weekly, monthly, or quarterly. A trend in time series is a tendency that shows the movement of a series to gradually higher or lower values over time, and it is frequently observed when the series has a rising or falling slope. Considering the Amazon stocks dataset, components of Open and Date columns are Multiplicatively and Additively decomposed.

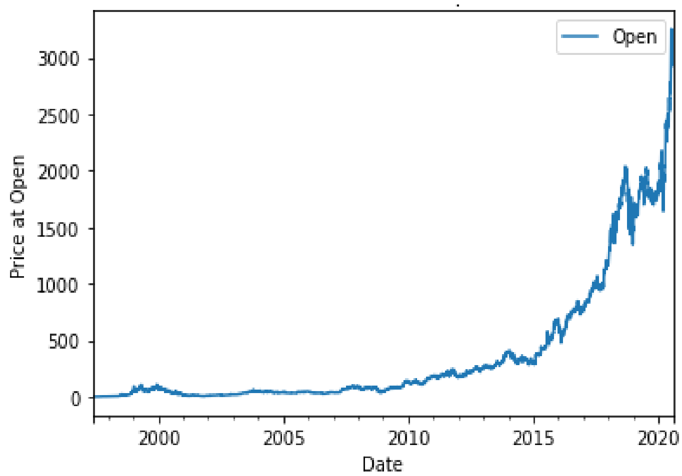

Figure 1: Distribution of Amazon Stocks when Market is Open.

3.1.1. Multiplicative Model

The error component of this model is multiplied by the trend and seasonal components before being added. From the formula **(t)=St x Tt x Rt,** Where,

St = seasonal component, Tt = trend-cycle component and Rt = remainder component

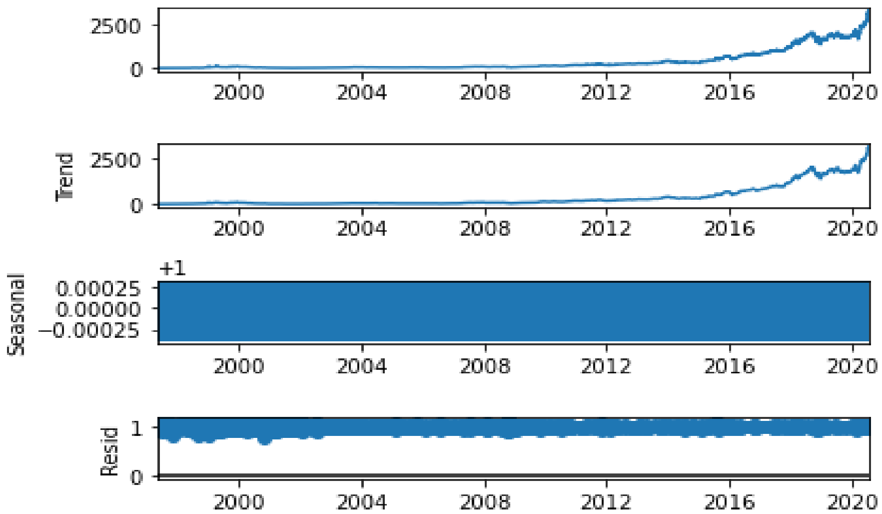

Figure 2: Multiplicative Model on Decomposed Series.

We can interpret from Figure 2 that the visualisations that are there shows a positive trend and seasonal looks like a blue blob and residual is showing high variability

3.1.2 Additive Model

The systematic component in this model is the arithmetic sum of the predictor's individual effects, and the variance of data does not change throughout different values of the time series. From the formula, **(t)=St +T t + Rt,** where

St = seasonal component, Tt = trend-cycle component and Rt = remainder component.

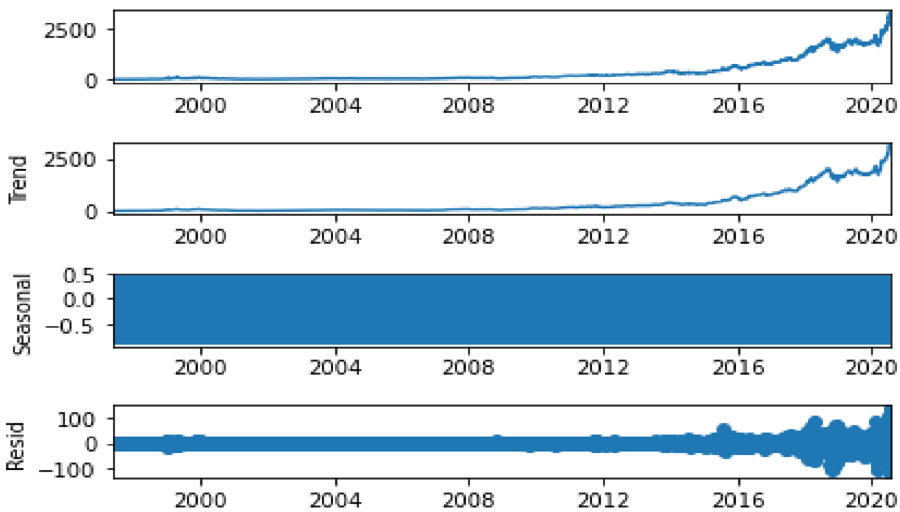

Figure 3: Additive Decomposed Data.

Similarly for additive model the trend is positive, and the season is also present, and we can clearly see the variability of the residual.

3.2 Importance of Augmented Dickey Fuller Test (Adf Test)

The ADF Test (Augmented Dickey Fuller Test) is a popular statistical test for determining if a time series is stable. It examines the null hypothesis that a unit root exists in a time series sample. The alternative hypothesis is typically stationarity or trend-stationarity, depending on the version of the test used. It's an improved version of the Dickey–Fuller test for a broader and more complex collection of time series models. The observations in a stationary time series are not dependent on time. Also, it is easier to model the time series when it is stationary. So, to check the stationarity of the data we used the adfuller () method in stats models. Ts a. Stat tools which offers a valid implementation of the ADF test in the stats model package. The null hypothesis assumes the presence of the unit root, that is, =1. To reject the null hypothesis, the p-value produced must be less than the significance threshold=0.0. As a result, we may conclude that the series is stationary.

After applying the AD Fuller test, we noticed that statistical value is greater than the significance value which implies that our data is accepting null hypothesis and it is not stationary.

To make the data stationary, the data will undergo first differencing by subtracting the shifted set and then re-plotting it.

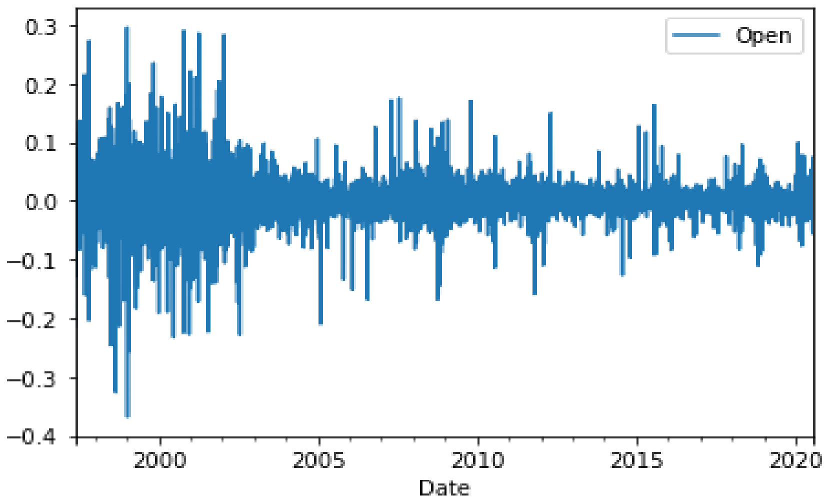

Figure 4: Series showing that Centra Tendency Components are not Changed with Respect to Time.

Now the statistical value is below significance value and our data is stationary**.** We also interpreted that the time series became stationary on first difference. Therefore, the number of differencing(d) of ARIMA and SARIMAX models will be 1.

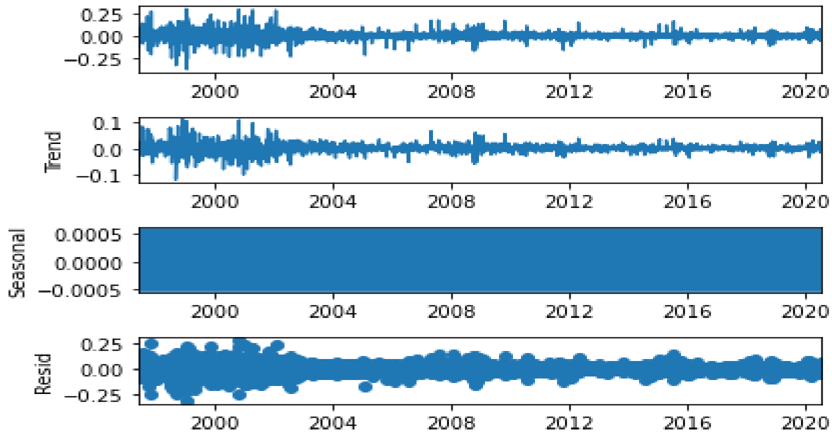

Figure 5: Decomposed Data.

ARIMA and SARIMAX models are characterised by 3 notations: p, q and d**.** The number of lag observations in the model, also known as the lag order, is denoted by the letter p. The number of times the raw observations are differed, also known as the degree of differencing, is defined by d. The size of the moving average window, also known as the order of the moving average, is denoted by the letter q.

3.3 Implementation of Arima

For time series forecasting, the ARIMA model is a well-known and widely used statistical technique. ARIMA (Auto Regressive Integrated Moving Average) is a technique that uses time series data to better comprehend a data set or predict future trends. The abbreviation is descriptive, expressing the key features of the model. They are, AR Autoregressive: A model that makes advantage of the dependent connection between an observation and a set of lagged data. I Integrated: It represents differencing of raw observations to make the model stable. MA Moving Average: A model that applies to lagged data the dependence between an observation and a residual error from a moving average model.

ARIMA forecasting is accomplished by importing time series data for the variable of interest. Statistical software will then determine the right number of delays or amount of differencing to apply to the data, as well as check for stationarity. It will produce the findings, which are frequently interpreted in the same way as a multiple linear regression model.

We have used the stats models package which allows us to fit an ARIMA model. After fitting the ARIMA model to the Amazon stocks Open rate dataset we have reviewed the residual errors. We have used an ARIMA (5,1,5) model, that specifies the lag value for auto regression(p) to 5, employs a difference order(d) of 1, and employs a moving average model(q) of 5. Likewise for p and q values we plotted PACF (Partial Autocorrelation) and ACF (Autocorrelation) graphs. These representations graphically depict the strength of a link between an observation in a time series and observations at previous time steps.

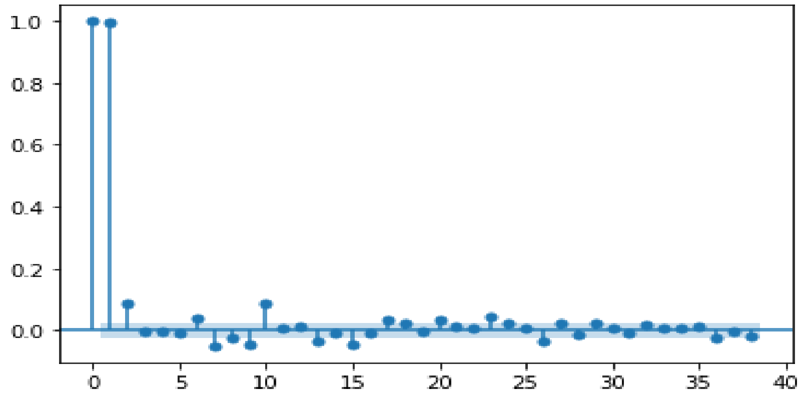

Figure 6: Partial Autocorrelation Function.

From this plot, 5 seemed to be the good starting point for the AR parameter (p) for the model.

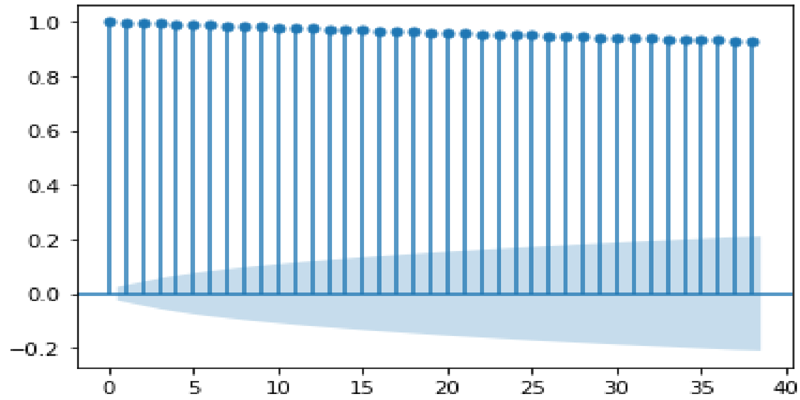

Figure 7: Autocorrelation Function.

From the autocorrelation graph we can say 5 to fit perfect for MA parameter (q).

ARIMA Model Results

Dep. Variable:	D.Open	**No. Observations:**	5450
Model:	ARIMA(5, 1, 5)	**Log Likelihood**	-17858.154
Method:	css-mle	**S.D. of innovations**	6.403
Date:	Mon, 10 Jan 2022	**AIC**	35740.308
Time:	00:43:07	**BIC**	35819.548
Sample:	05-16-1997	**HQIC**	35767.960
	- 04-05-2018		

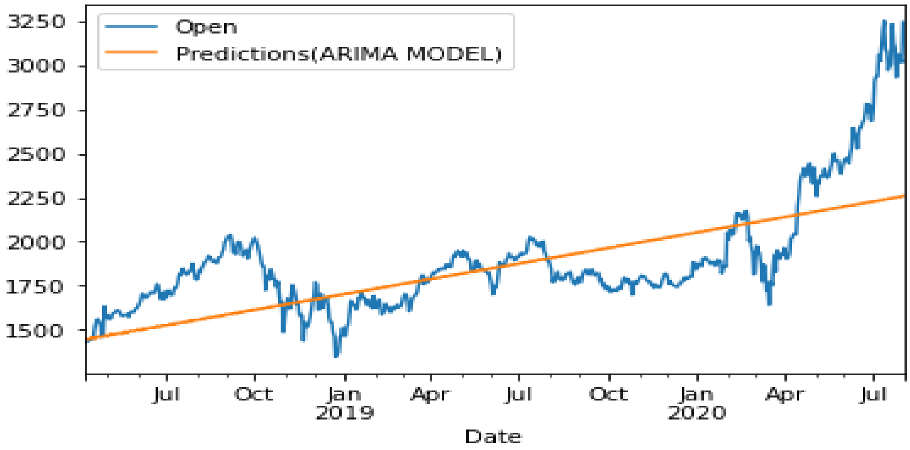

Figure 8: Amazon Stock Forecast by ARIMA MODEL and its Corresponding Statistical Results.

After building and fitting the ARIMA (5,1,5) model we trained and tested the data. 90% data is taken for training and remaining for testing set. Then the predicted ARIMA model is plotted.

From the above graph we can understand that there are variations in stock prices throughout the year. The prices shoot up in the month of October but there's a fall during January. Then for the year 2020, the stocks prices increase rapidly after April.

3.4 Identification of Seasonal Pattern using SARIMAX

SARIMAX stands for Seasonal Auto-Regressive Integrated Moving Average with eXogenous Factors and is an extension of the ARIMA family of models. It's a seasonal equivalent model, similar to SARIMA and Auto ARIMA, that can handle external influences. Abbreviation for SARIMA model is **SARIMA (p, d,q) * (P,D,Q,S)** where p, d, q are for non-seasonal parameters and P, D,Q,S stands for seasonal parameters where S is length of repeating seasonal pattern. We have used the same **stats models package** to fit the SARIMAX model. Using **SARIMAX (5,1,5) * (5,1,5,12) model** the results are calculated and the prediction graph is plotted. The equation on which SARIMAX for a univariate structural model can be represented as

$$y_t = u_t + \eta_t$$
$$\phi_p(L)\tilde{\phi}_P(L^s)\Delta^d\Delta_s^D u_t = A(t) + \theta_q(L)\tilde{\theta}_Q(L^s)\zeta_t$$

The SARIMAX model requires the identification of p, d, and q standards to be done using the auto co-relation function and partial auto co-relation function.

SARIMAX Results

Dep. Variable:	Open	**No. Observations:**	5451
Model:	SARIMAX(5, 1, 5)x(5, 1, 5, 12)	**Log Likelihood**	-17838.039
Date:	Mon, 10 Jan 2022	**AIC**	35718.077
Time:	18:13:07	**BIC**	35856.702
Sample:	05-15-1997	**HQIC**	35766.457
	- 04-05-2018		
Covariance Type:	opg		

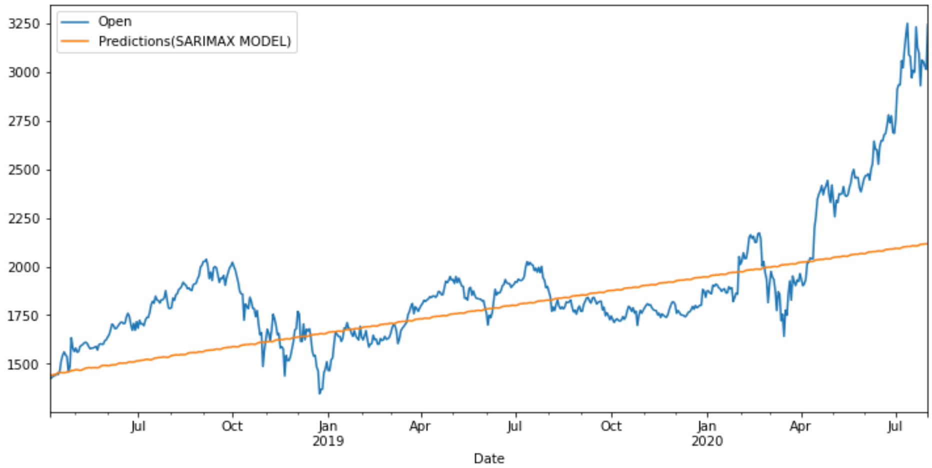

Figure 9: Amazon Stock Forecast by SARIMAX Model and its Corresponding Statistical Results.

From the Figure 9 above we can understand that there are variations in stock prices throughout the year. Prices went up in October but there was a fall in mid-January. Then in 2020, stock prices rose sharply after April.

4. RESULTS & DISCUSSIONS

Prediction is the process of creating predictions based on previous and current data. Stock analysts utilise a variety of forecasting methodologies to analyse the quantity of future stock movements. Predictability also provides an important level of organization. A data view is an image of numerical data. After anticipating stock market trends, we use line charts, candlestick charts, charts, and histograms to display the short-term consequences of investing aid. The x-axis depicts the passage of time in years/months, while the y axis depicts stock values. Seasonal and non-seasonal features refer to the nearest distance or area away from the actual graph price line. The different metrics used here are RMSE, MAE and R^2 Value. The amazon stocks don't have a defined seasonality pattern and thus we compared ARIMA and SARIMAX model. The SARIMAX model outperforms the ARIMA model just by 0.1% in the field of all the three-accuracy metrics. The R^2 Value comes very near to 0 thus the model is comparatively believed to be well sufficient to predict the cost of AMAZON stocks for the current next decade.

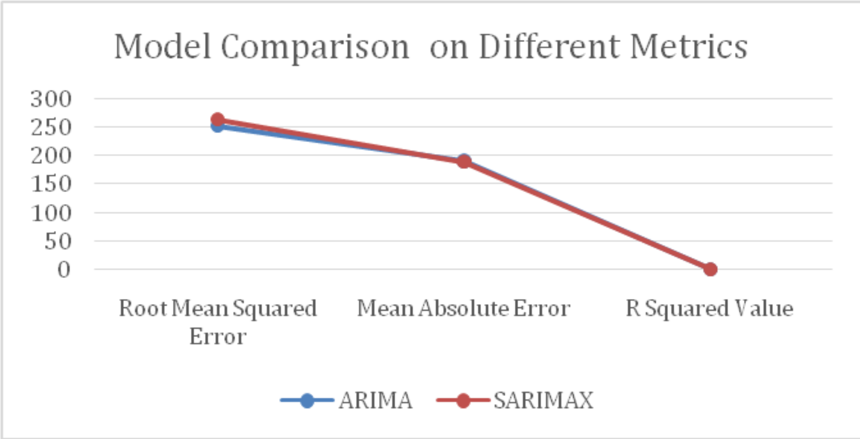

Figure 10: Model Comparison on Different Metrics.

5. CONCLUSIONS

Stock market forecasts are essential for a successful firm. Forecasts are usually beneficial in lowering the risk element in any company situation. Historical data and prior company trends may be used to assess the risk element. In this research paper we have predicted future forecasting of the cost of Amazon Stock by implementing ARIMA model and SARIMAX model. In the real data set we have only considered date and open column then we have decomposed the data by multiplicative decomposition and additive seasonal decomposition from which we conclude that the trend is positive. Then we applied AD Fuller test to check stationarity. After first differencing the statistical value is below significance and our data was became stationary. We then trained the ARIMA MODEL by give the value of AR lag(p), MA lag(q) and differencing term(d). After implementing ARIMA we found seasonal pattern which is overall increasing. We have also implemented and trained the SARIMAX MODEL where we have predicted the future forecasting which is overall increasing. Test results obtained demonstrated the strength of the SARIMAX predictive model. On a short-term basis, stock price indexes. This can help stock market investors make a good investment choice to purchase, sell, or hold a stock. The obtained ARIMA model can compete well with short-term forecasting prediction tactics based on the outcomes.

6. REFERENCES

1. *Tae Kyun Lee et al. "Global stock market investment strategies based on financial network indicators using machine learning techniques", Expert Systems with Applications, 117(2019):228-242.*

2. *Bruno et al. "Literature review: Machine learning techniques applied to financial market prediction", Expert Systems with Applications, 124(2019): 226-251.*

3. *Thomas Fischer et al. "Deep learning with long short-term memory networks for financial market predictions", European Journal of Operational Research, 270.2(2018): 654-669.*

4. *Kang Zhang et al. "Stock Market Prediction Based on Generative Adversarial Network", Procardia Computer Science, 147(2019):400-406*

5. *Tao Xing, Yuan Sun, Qian Wang, Guo Yu, "The analysis and prediction of stock prices", IEEE International Conference on Granular Computing (GrC), December 2013, pp. 368–373.*

6. *Vishwanath R. Ha, Leena Sa, Srikantaiah K. Ca, K. Shreekrishna Kumar b., P. Deepa Shenoya, Venugopal K. Ra, S. S. Iyengarc, L. M. Patnaik, "Forecasting stock time-series using data approximation and pattern sequence similarity", International Journal of Information Processing (IJIP), September 2013, pp. 90-100.*
7. *Ayodele A. Adebiyi, Aderemi O. Adewumi, Charles K. Ayo, "Stock price prediction using the ARIMA model", 16th IEEE International Conference on Computer Modelling and Simulation (UKSim), March 2014, pp. 106 -112.*
8. *Roy, Samanwita. "Comparative Flow Analysis of NACA S6061 and NACA 4415 Aerofoil by Computational Fluid Dynamics." International Journal of Mechanical Engineering (IJME) 7.2 (2018): 9 18.*
9. *Pittala, Suresh, and T. Diriba. "Computational Fluid Dynamics Analysis of Impeller Design For A Pump." International Journal of Mechanical Engineering (IJME) 5.4.*
10. *Padmavathi, M., and Shaik Mahaboob Basha. "A Conceptual Analysis on Cloud Computing ACO Load Balancing Techniques." International Journal of Computer Science and Engineering (IJCSE) 6.4 (2017): 39-48.*